Assessment literacy and the mathematics teacher

Cornelia C. Soto, Ateneo de Manila University, Philippines

Assessment

Assessment has always been a part of the teaching-learning process but its role within the larger picture of the instructional process has always been distorted (Black in Klenowski & Wyatt-Smith, 2014). For students and parents, classroom assessment has been equated with tests and examinations. These tests and examinations with their selected response and supply formats have resulted in negative reactions and unpleasant memories on students and parents. To a large extent, these anxiety-producing events (quizzes, long tests, exams) and formats (multiple choice, true or false, matching type, short response) are still around. Fortunately, there are ways to make these formats and events less intimidating, more useful and more powerful. Today, there are many other ways of assessing knowledge and skills and many other ways of getting information on students' understandings than tests and exams. Classroom assessment at present is quite different from the common image of classroom assessment comprised mostly of testing and examinations.

For teacher practitioners, assessment has come to be viewed as the negative aspect of the teaching-learning process. This is especially so when administrators examine and use the results of assessment to hold teachers accountable for student learning. Aside from this, assessment results are also often used as the rationale for raising the standards of student achievement. When assessments are used as a basis, sometimes the sole basis, for accountability and higher standards, teachers tend to teach-to-the test or teach-the-test. In this sense, assessment drives instruction. Teaching approaches tend to focus on transmission of knowledge. Questions are formulated to simulate questions and items that may appear in the tests. Consequently, more premium is given to questions that are precise, require accurate answers and can easily be categorized as correct or incorrect.

Assessments used for purposes of accountability and standards are thus, high-stakes tests. Studies show the negative impact of these high-stakes tests on the following: students' motivation for learning, adopting a teaching style that emphasizes transmission of knowledge, constricting effects on the curriculum, more time for practice tests, less time on other student performances. (ARG, 2002a)

When tests alone are used to measure learning outcomes, there is a narrow and naïve view of learning. There is a lack of match between the education goal of preparing students for life and what was being taught as dictated by tests. It is an accepted belief that what is needed is a curriculum designed "to develop the ability to access and evaluate information, to apply knowledge to new situations and acquire the prerequisites for learning throughout life" (Harlen, 2007, p2).

In the last few decades, the amount of assessment, the variety of forms and the uses of assessment have increased dramatically (Frey, 2014). Millions of unique assessments are constructed yearly (Worthen, Borg, & White, 1993). A surge of formative assessment systems has been seen in different forms and variations (Arielli-Attari, 2012). As a consequence, there is now a need for explicit training for teachers in the new approaches, understandings and best practices in the areas of classroom assessment.

If teachers are to engage effectively in the different aspects of instruction, they need to be assessment literate.

Classroom Assessment
Classroom assessment is a systematic collection of information about students' abilities, characteristics, skills, understanding and knowledge developed, administered and scored by a teacher for the purpose of evaluation (Frey, 2014). Classroom assessment is a broad set of classroom activities done for the purpose of gathering information and evaluating students' learning. It is used to assess not only the students but the teachers as well. It can be done before instruction, during instruction and after instruction. Information gathered during assessment can be used by teachers to improve instruction or to summarize student achievement and by students to monitor their learning. When assessment results are used to inform instruction, assessment has a developmental function. Assessment becomes assessment **for** learning or formative assessment. When assessment summarizes student achievement or what students have learned and how much students have learned, assessment has an accountability function. It becomes assessment **of** learning or summative assessment. When students use the results of assessment to monitor their learning or to improve their learning, assessment becomes assessment **as** learning, also called formative assessment.

Properties of assessment
Two foundational ideas, validity and reliability characterize assessment. Validity is the characteristic of the assessment when it measures what it is supposed to measure. Validity has been described as the most "fundamental consideration in developing and evaluating tests" and is defined as the "degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (AERA, APA, & NCME, 1999, p 9). There are different types of evidences that can be examined to justify that an assessment is valid: content-based validity, criterion-based validity and construct-based validity. Reliability on the other hand, refers to the extent to which the assessment if repeated would give the same results. Assessment is said to be reliable when the assessment results are of acceptable consistency and accuracy.

Aside from validity and reliability, two other properties of assessment can be taken into account: impact on other assessment practices and the use of resources (Harlen, p 21). Impact means the consequences of the assessment in relation to the uses that are made because of its results. As mentioned at the beginning, assessment has always been a part of the teaching-learning process so it may be important to mitigate any adverse effects of assessment on the school's stakeholders. For instance, when there are external tests, teachers administer more summative assessments than needed.

Resources required in an assessment may be human resources such as teacher's time and capabilities or material resources such as cost, paper, and the Internet. These resources must be commensurate to the value of the assessment results to the users of the information.

Current Perspectives on Assessment
There are several perspectives on classroom assessment that are presently being investigated by researchers and being used by teachers: summative assessment or traditional paper-and-pencil assessment, formative assessment, performance-based

assessment, authentic assessment, standards-based assessment and universally designed assessment.

Summative assessment
Summative assessment or assessment of learning is still the most common approach to classroom assessment. That is, assessment is done after instruction, measures what students have learned, with formal test-taking rules and scoring procedures meant to determine grades of students, and information gathered basically for teacher's use. Summative assessment is usually in the form of selected response or supply type of questions. Selected response type includes true-or-false, matching exercises, and multiple-choice. Supply type or constructed response type includes short response and essay. These formats are the usual paper-and-pencil tests students and parents have come to expect. Traditional paper-and-pencil tests are still very popular and considered efficient since student achievement can be easily and reliably measured. In these tests, the students' answers to questions are assumed to be evidences of their learning. The concern here then is on the validity of the assessment results. To establish validity, the match between the instructional objective and the assessment results is examined. In selected response types where students "can" guess the answer, the issue is reliability.

Summative assessment results are used for individual students and cohorts of students. For individual students, the uses may be internal or external (Harlen, 2007 p16). Record keeping, giving grades, reporting to students and parents, informing students about consequent courses to take are some internal uses of summative data. Certification for vocational qualifications, selection for higher education and licensure are sample external uses of summative results. For groups of students, summative data are used for accountability and monitoring. Summative data are used for accountability when data are used to evaluate teachers, administrators and schools.

In mathematics, traditional paper-and-pencil test formats can be categorized into basic response types and complex response types. Under the basic response types are selected response types that include multiple choice, multiple selection-multiple choice while the constructed response types include short response which may be numeric entry, mathematical equations and expressions, kinds of graphs and short answers. The term basic does not imply easy but rather means direct or straightforward as opposed to a complex response type (Graf, 2009).

Complex response types include responses that present argument or justification as in proofs, solution to a problem, or a variety of representations such as equations, diagrams, lists, charts and graphs.

Performance-based assessment
Performance-based assessment is an approach to assessment where students demonstrate their abilities and skills either through a performance or the construction of a product. This approach started to be popular about three decades ago and is now regarded almost as traditional as paper-and-pencil tests. It is considered the best format for assessing skill or ability and has led to the use of new scoring options – the scoring rubric. The scoring rubric contains the important components of the performance task and is related to the purpose of the assessment and its observable parts or steps. Scoring is based on either ratings or rankings.

In the early 1980s, three teachers wanted to show the literacy skills of their students in the Primary Language Record, a collection of students' writings and other products. Called "portfolios" after the large folders carried by artists to show their work, these individualized collections were seen as a more meaningful alternative to a single score or grade. Eventually, the use of portfolios spread to other populations and other disciplines. Frey (2014, p. 165) summarized the features of portfolios as person-centered, qualitative and holistic; usually accentuate the positive, involve the student; have a rich variety of student product over time; involve reflective self-evaluation. The portfolios show what students know and can do and how those skills develop over time. Portfolios also make multi-modal presentations of products possible so that visual, audial and kinesthetic modes can combine with digital applications for presentation and storage purposes.

Formative assessment
Formative assessment is an assessment approach used by teachers "to provide feedback to students to enhance their motivation and learning by designing instruction to meet students' needs" (McMillan, 2007, p. 1). Comprehensive reviews and subsequent research show that "the use of effective formative assessment strategies enhances student achievement, whether measured by classroom or large-scale tests" (Black & William, 1998; Black & William, 2005; Brookhart, 2005 as cited in McMillan, 2007).

In using formative assessment, teachers utilize a cycle of evaluating student work and behavior, giving feedback to students and applying instructional activities (see Figure 1). When students are given feedback during a lesson before instruction is completed, students can monitor and improve their own learning and academic progress. Teachers can use information gathered during instruction to modify their teaching or use an alternate strategy while instruction is still ongoing.

In formative assessment, feedback and clear criteria (Brookhart, 2014) are the key components that facilitate improved learning. Feedback is effective when it is immediate and specific. If feedback is given after a time lag, it will no longer be meaningful to the student who has moved forward to another task. Feedback must also be specific and contingent to the task (clear criteria). Telling a student to "study harder" is not feedback because s/he would not know how to study harder. Similarly, comments like "very good," "excellent," or "poor" cannot be considered formative feedback because students do not know why their work is "good" or "excellent" or " poor" and will therefore not know how to proceed. Feedback to be formative must contain a procedure or a formula on what specific action to take. When specific feedback is given, the student can determine where s/he is in his/her learning; establish where s/he needs to go and determine what s/he needs to do to get there (Ramaprasad, 1983, in McMillan, 2007, p. 32). This allows students to monitor and control their own learning.

The main goals of learning in the mathematics classroom are problem solving, communicating, and reasoning (Harlen, in McMillan (Ed.), 2007, p 117). Problem solving involves development of skills for making decisions about which mathematical knowledge to use and decisions about how to use this knowledge when confronted with a problem. Problem solving also involves progressively greater autonomy and more confidence in the approach to mathematical problem situations. The communication

strand tackles the issues of students communicating the mathematics they are doing (sending the message of what they understand and how much they understand) and understanding the mathematics presented to them (receiving the message given to them). The reasoning strand involves justifying, generalizing, and ultimately, proving mathematical results. These skills are usually taken to include information-processing, reasoning, creative thinking, evaluation and inquiry skills. The mathematics teacher who wishes to use formative assessment must design his/her formative assessment in such a way that s/he can gather information about student thinking and understanding when they solve a problem, communicate about the math they are doing or reason about their math results.
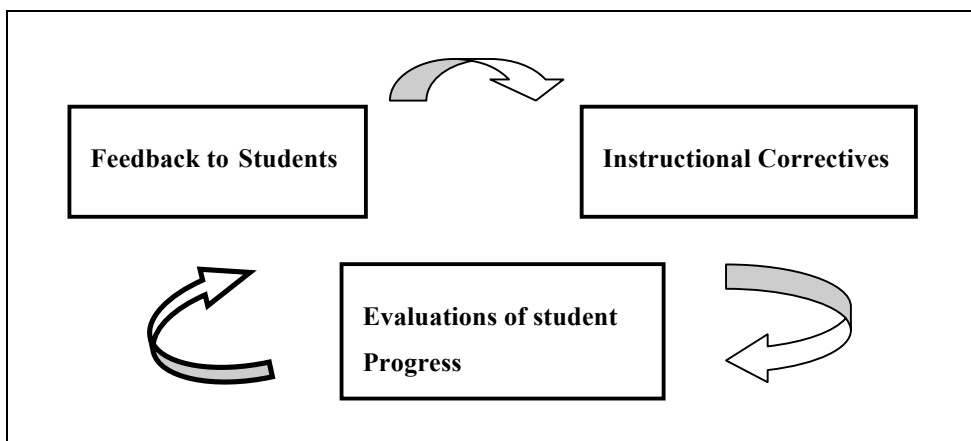

Figure 1. J. McMillan's (2007) The Formative Assessment Cycle

An approach to gather evidence of students' thinking and understanding is through the use of effective questioning.

Watson and Mason (1998 as cited in Harlen, 2007) give sample questions in mathematics that are not limited to students' knowledge or recall of facts but can show what they know about a situation and its general properties. For instance, instead of asking "Is 7 a factor of 21?" more information on students' thinking will be demonstrated if the question becomes "Why do you think 7 is a factor of 21?" The latter question asks students to explain and justify their reasoning while the former question is simply a recall or a retrieval question. The following questions provide teachers with more information about students' thinking and understanding while asking students for deeper thinking or extended thinking:
  • Is it always true that…?
  • Is it ever false that…?
  • How many different ways can you find to…?
  • Give me a definition of… in your own words.
  • What is the same and what is different about…?
  • What is wrong with the statement…?
  • What other information do you need to solve this problem?

Frey (2014, pp. 79-80) enumerates formal formative assessment that include quizzes or exams that do not impact on students' grades; conferences where teacher and student discuss work plans and strategies; practice exercises in workbooks and textbooks; self-scoring rubrics for quality skills and abilities; checklists with steps or objectives on a scale (e.g. almost there, not yet, etc.); self-reflection worksheets to identify areas of strengths and challenges. These formative assessments show that the format of the assessment can be similar but the purpose or function of the assessment instrument is what helps determine whether assessment is summative or formative.

Authentic assessment
Authentic assessment is assessment that matches tasks and expectations in the real world. Frey (2014, p. 203) describes assessment as authentic when the "tasks, content, expectations, and evaluation methods of the assessment are similar to the meaningful tasks, content, expectations, and evaluation methods outside the classroom in the real world. Authenticity differs according to the developmental stage and context of the learner. It may be playing and socializing, technology, democracy, tertiary education or the real world of work. Since these real-world expectations are valued outside of the classroom, the assessment is meaningful and becomes more motivating for students. Authentic assessment also addresses educators' concerns over preparing students for the realities of the real world.

In mathematics, it may be more difficult to define what an authentic task is because of the apparent discrepancy between what happens in the mathematics classroom and what real mathematicians in the real world do. For instance, problem solving in the mathematics classroom is usually a solitary activity while in the real world, problem solving as done by mathematicians is collaborative. Goos (cited in Frey, 2014, p. 218) suggests that authenticity in mathematics can focus on what is "valued in the real world." Then, the mathematical processes of problem solving, making connections and communication may be considered authentic.

Standards-based assessment, judgment and moderation
In education, standards are defined and understood as "fixed points of reference for assessing individual students and cohorts" (Sadler, 1987 as cited in Klenowski & Wyatt-Smith, 2014). Standards are fixed points so that teacher judgments may be consistent. However, in practice, interpretations of the "fixed point" remain subject to individual differences. Hence, there is a need for moderation. Moderation is described as a process involving teachers in discussion and debate about their interpretations of the quality of assessed work (Linn, 1993 in Klenowski & Wyatt-smith, 2014). It is only when standards are defined and applied, deliberated and debated on that standards can "become published indexes or features of quality against which judgment can be made available for scrutiny, and thereby made defensible" (Klenowski & Wyatt-Smith, 2014, p. 13).

Standards-based assessment is thus, an assessment approach that focuses on characteristics of quality evidenced on the assessed work of students. It asks teachers to apply their expertise and knowledge to make professional judgments on the quality of student work that is being assessed. It asks teachers to express their personal interpretations of the standards with the aim of arriving at a consensus on assigning a grade or standard on the work under study.

Universally designed assessment
Universally designed assessment is assessment that emphasizes accessibility and fairness for all students regardless of gender, ethnicity, mother tongue, disability, race. The concept of universal design can be traced back to the field of architecture and engineering in the 70s when architect Ron Mace developed the building accessibility code in North Carolina. Buildings and physical environments started to be designed such that they were accessible to all, including those in wheelchairs (Bowe, 2000 as cited in Frey, 2014).

The approach has spread to include other populations and other areas, including education. From standards for the physical environment, the standards have been applied to processes which include physical and experiential aspects.

There are seven established standards for universal design (Center for Universal Design, North Carolina State University, 1997):
1. Equitable use - The design is useful to people with diverse needs and abilities.
2. Flexibility in use - The design accommodates diverse preferences and abilities.
3. Simple and intuitive use - Use of the design is easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level.
4. Perceptible information - The design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities.
5. Tolerance for error - The design minimizes hazards and the adverse consequences of accidental or unintended actions.
6. Low physical effort - The design can be used efficiently and comfortably and with a minimum of fatigue.
7. Size and space for approach and use - Appropriate size and space is provided for approach, reach, manipulation, and use regardless of user's body size, posture, or mobility.

Suggestions and guidelines for universally designed tests are found on http://www2.lexcs.org/osep/pdf/Universal_Design_LSA.pdf and include suggestions on:
1. bubble answer options - fewer errors if students answer on the test itself
2. contrast - off-white or light pastel, non-glossy paper, type should be black
3. graphs and tables - black and white, symbols on the graph or chart instead of a legend somewhere else
4. illustrations - black-and-white is better, avoid green and red if colored, omit if irrelevant to the task
5. leading - amount of space between lines, 2-4 points
6. text format - left justified, unjustified or "ragged" on the right, better to use standard typeface with upper and lowercase letters than all uppercase or italicized texts, bolding is better than all caps.
7. type size or font size - 14 point type is better than 12 or 10, for students with moderate visual impairment, point 18; fonts that are serif, that is, with perpendicular lines at the end of each stroke such as Times New Roman or Courier may work better than sans serif (e.g. Arial or Helvetica)
8. white space - also called blank space, half the page should be blank, more white space for younger students

Conclusion

Assessment has become a "hot" topic in education. Teachers today can choose from different approaches to classroom assessment either to help improve learning or to summarize what has been learned. Assessment is not simply a tool to measure learning but also a process or a combination of tool and process to facilitate learning by all kinds of students.

References

AERA, APA, & NCME (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

ARG (Assessment Reform Group) (2002a). Testing, Motivation and Learning. (Obtainable from the ARG website: www.assessment-reform-group.org)

Arielli-Attali, M. (2012). *Formative Assessment with cognition in mind: The cognitively based assessment of, for and as learning (CBAL) research initiative at Educational Testing Service.*

Brookhart, S. M. (2014). *Expanding views about formative classroom assessment: A review of the literature.* In J. McMillan (Ed.), *Formative classroom assessment: Theory into practice.* New York: Teachers College Press, Columbia University.

Elliott, J.L., Thurlow, M.L., & Ysseldyke, J.E. (1996). *Assessment guidelines that maximize the participation of students with disabilities in large-scale assessments: Characteristics and considerations* (Synthesis Report No. 25). Minneapolis, MN: University of Minnesota. Retrieved on March 25, 2015 http://education.umn.edu/NCEO/OnlinePubs/Synthesis25.html.

Frey, B. B. (2014). *Modern classroom assessment*. Thousand Oaks, California: Sage Publications, Inc.

Graf, E. A. (2009). *Defining mathematics competency in the service of cognitively based assessment for grades 6 through 8*. (Research Report No. RR-09-42). Princeton, NJ: Educational Testing Service.

Harlen, W. (2007). *Assessment of Learning*. Thousand Oaks, California: Sage Publications, Inc.

Klenowski, V. & Wyatt-Smith, C. (2014). *Assessment for Education: Standards, Judgment and Moderation,* Thousand Oaks, California: Sage Publications, Inc.

McMillan, J. H. (2007). *Formative classroom assessment: Theory into practice.* New York: Teachers College Press, Columbia University.

Thompson, S. & Thurlow, M. (2002). *Universally designed assessments: Better tests for everyone!* (Policy Directions No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved on March 25, 2015 http://education.umn.edu/NCEO/OnlinePubs/Policy14.htm.

Worthen, B.R., Borg, W.R., & White, K.R.(1993). *Measurement and evaluation in the schools.* White Plains, NY: Longman

_____

Cornelia C Soto
Ateneo de Manila University, Katipunan Avenue, Loyola Heights, Quezon City, 1108 Philippines
csoto@ateneo.edu